SVM and MSVM for plant disease detection: A Comparative Study

Rajesh Kumar
Department of Computer Science and Engineering
Chaudhary Devi Lal University
Sirsa, India
rajesh31070@gmail.com

Vikram Singh
Department of Computer Science and Engineering
Chaudhary Devi Lal University
Sirsa, India
vikramsinghkuk@yahoo.com

Abstract*— This study performs a comparative analysis of Support Vector Machine (SVM) and Multi-Class Support Vector Machine (MSVM) for plant disease detection using a dataset consisting of images of healthy and diseased plant leaves. Feature extraction is performed using a gray level co-occurrence matrix (GLCM) and the resulting features are then used to train both classifiers. SVM is used for binary classification to distinguish healthy and diseased plants, while MSVM is used for the multi - class classification of various plant diseases. The results reveal that both classifiers have high accuracy where MSVM exceeded the classification of several diseases. This study highlights the effectiveness of machine learning methods for plant disease detection, achieving 90% accuracy for the MSVM model and 89% accuracy for the SVM model.

Keywords*— SVM, MSVM, disease detection, GLCM, binary classification.

I. INTRODUCTION

Herbal disease testing is the main area of research that can significantly affect agricultural productivity and food safety. Early and accurate identification of plant diseases can be intervened in time to reduce yield losses and improve yield quality. Traditional methods of disease detection, often relying on expert knowledge and manual inspection, can be labor intensive and prone to inaccuracies. Recent advances in machine learning (ML) have offered promising alternatives, particularly with the application of Support Vector Machines (SVM) and its extensions.

Support vector machines (SVMs) are widely used in classification tasks due to their robustness and efficiency in high-dimensional spaces. SVM works by finding the hyperplane that best separates different classes in the feature space. In connection with several categories (such as plant disease test), standard SVM must be changed. SVM on multiple levels (expansion of SVM frames) adopts strategies such as a VS-One (OVO) and one-VS-Rest (OVR) to effectively handle multiple classes.

This research aims to conduct a comparative analysis of standard SVM and Multi-SVM approaches in the context of plant disease detection. By evaluating their performance across various metrics, the study also delves into understanding their relative strengths and limitations. The objectives of the research are:
To evaluate and compare the performance of standard SVM and Multi-SVM in the context of plant disease detection.
To analyze the effectiveness of different Multi-SVM strategies, such as one-vs-one and one-vs-rest, in handling multi-class classification problems.
To assess the accuracy and performance, via metrics of SVM and Multi-SVM models in detecting plant diseases.

Related Work
Recent advancements in deep learning frameworks have significantly improved plant disease detection, particularly for diseases causing changes in leaf size, shape, and brightness. Authors in [1] have demonstrated that these frameworks offer highly accurate keypoint extraction, which enhances the ability to differentiate between various plant diseases and is effective with datasets like PlantVillage. Compared to other methods, [2] researchers have used U-Net and starting net models to accurately determine the condition of the leaves. The study [3] emphasized the efficiency of learning (OSL) technology in groups of plant diseases. It enters real-time applications by enabling fast processing and reducing the need for large amounts of labeled data.

The integration of Convolutional Block Attention Module (CBAM) with pre-trained Convolutional Neural Networks (CNN) has also been shown to improve classification accuracy, especially in lightweight models such as EfficientNetB0, as demonstrated by [4] Remarkable performance is achieved with limited training data. Researchers [5] showed that hybrid fuzzy C-segmentation combined with deep learning classifiers improved the detection of banana leaf diseases such as black leaf spot and bacterial wilt, highlighting the importance of accurate segmentation.

A study [6] combined Ant Colony Optimization (ACO) with CNN to achieve superior foliar disease identification. Researchers [7] indicated that using machine learning techniques such as Random Forest Classifier (RF) for wheat disease classification achieved high accuracy. In a study [8], CNN showed better performance than Support Vector Machine (SVM) and K Nearest Neighbor (KNN) in leaf disease detection. Microspectral imaging including SVM and KNN, has been utilized to observe subtle changes in leaf color and water content, leading to improved disease classification in [9]. Lab color space mapping and random sample consensus (RANSAC) for curve fitting have enhanced feature extraction for leaf disease identification across benchmark datasets in [10].

A recent evaluation of the Vision Transformer (ViT) model showed better performance compared to traditional CNN architectures, especially for large datasets, despite slower prediction times [11]. A hybrid approach combining ViT with CNN aims to balance accuracy and speed. The AFD-Net model has demonstrated remarkable performance in classifying apple foliar diseases, outperforming the validation accuracy of another architecture [12].

979-8-3315-1905-6/24/$31.00 ©2024 IEEE
DOI 10.1109/ICACCTech65084.2024.00033

Systematic analyses reveal that deep learning models, particularly CNNs and ResNet, consistently outperform traditional machine learning techniques in plant leaf disease classification [13]. Challenges remain, including model generalization and the need for standardized disease severity estimation methods. Noteworthy achievements include high accuracy rates for models such as ResNet-9 and ResNet-34-based Faster-RCNN, with advancements in hyperspectral remote sensing providing early plant disease detection capabilities, despite standardization challenges [14, 15]. DenseNet-77 models have proven effective in detecting tomato plant leaf diseases, even with variations in image quality [16].

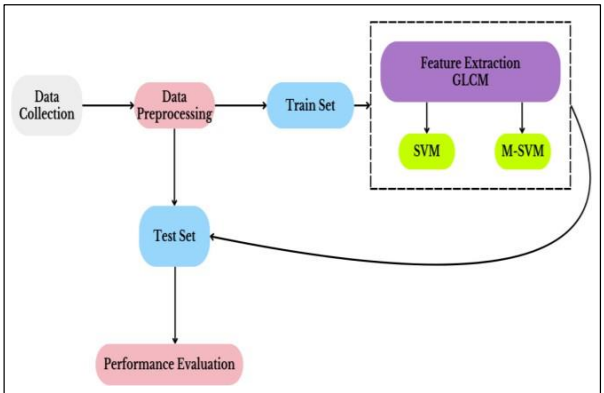

Fig 1. Proposed System Architecture

Fig. 1 demonstrates the proposed system used in our experimental analysis. It analyses the following:

- To collect the dataset comprising images of plants labeled as "healthy" or "diseased," with diseased images further categorized by specific diseases.
- To preprocess the dataset images by resizing them to 256x256 pixels, normalized to a range of 0 to 1, and augmented with rotation, flipping, and zooming.
- To extract features using GLCM.
- Model Implementation:
  - **SVM:** To classify images into "healthy" or "diseased."
  - **MSVM:** To implement a "one-vs.-all" approach for multiple plant diseases.
- To train the models with 80% of data and validate with 20%, with fine-tuning of parameters.
- To evaluate the performance of the system using F1-score, accuracy, precision, recall, and confusion matrices.

II. METHODOLOGY

This section summarises the steps followed in carrying out the experimentation delegated in Section II.

A. Dataset Collection

The dataset comprises web scraped images of plants collected from various farming areas. Each image is labeled as either "healthy" or "diseased," with diseased samples further categorized based on the specific type of disease. It includes images featuring leaves with visible signs of stress or damage, indicating disease, while the bottom row shows healthy, undamaged leaves. Table I demonstrates sample images from the two classes.

TABLE I. SAMPLES FROM THE DATASET

Class	Shuffled Sample 1	Shuffled Sample 2
Healthy		
Diseased		

B. Dataset Preprocessing

All images are tailored to a uniform size of 256x256 pixels to standardize the entire size of the model. The normalization of the pixel value is from 0 to 1, which helps to converge faster and improve the efficiency of model training. Use data enhancement techniques such as rotation, flip, and scaling to increase the robustness of the model and improve its generalizability. image. Figure 2 shows the subsequent preprocessing steps.

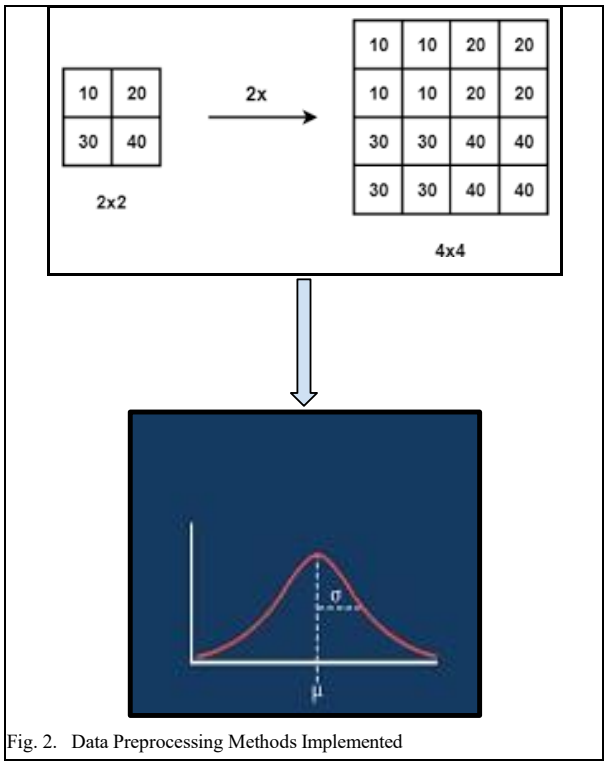

Fig. 2. Data Preprocessing Methods Implemented

C. Feature Extraction

GLCM is used for feature extraction. It identifies key features in input images. The second-to-last layer of the network is

utilized to extract high-level features from the input images, which serve as feature vectors for further classification tasks.

Let f(x) represent the feature vector output from the GLCM model for an input image x. Function vector F (X) is used as input for SVM and MSVM classification machines for plant diseases. Extracted features allow SVM and msvm to master the characteristic features of healthy and sick leaves that facilitate the exact classification

.

D. Models

SVM

The preprocessed images are split into groups of healthy and diseased classes using SVM [17]. The hyperplane that best divides the classes in the feature space is derived using:

$$f(x) = sgn(w^T x + b)$$

where w is the weight vector, x is the feature vector, b is the bias, and sgn is the sign function.

MSVM

MSVM uses the "one-vs.-all" method to generalise samples into the different classes. The winning class has the best decision score [18]. Here:

$$f_k(x) = w^T x + b_k$$

E. Model Training and Validation

The model training and validation process is performed using a split of 80% for training and 20% for validation. SVM and MSVM classifiers are trained on the extracted feature vectors from the GLCM model. SVM is used to classify images as "healthy" or "diseased", while MSVM handles the classification of different types of plant diseases into multiple classes. To eliminate redundancy and increase generalization, hyperparameters such as regularization parameters are adjusted during training.

F. Performance Evaluation

The following performance metrics are evaluated to establish the performance of the system.

Accuracy

Accuracy is the ratio of accurately predicted instances to the total number of cases, including True Positives and True Negatives.

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP}$$

Where, TN is True Negatives; TP is True Positives; FN is False Negatives; and FP is False Positives.

Precision

By comparing the percentage of accurately predicted positive instances to all anticipated positive instances, precision calculates the accuracy of positive predictions.

$$Precision = \frac{TP}{TP + FP}$$

Recall

Recall quantifies the model's ability to identify every instance of a positive. It is the percentage of all positively impacted occurrences that were correctly predicted to happen.

$$Recall = \frac{TP}{TP + FN}$$

F1

An impartial assessment of a model's performance is offered by the F1-Score, which is the harmonic mean of recall and precision. It provides beneficial information in cases where there is an uneven distribution of courses or data points.

$$F1\ Score = \frac{2 * Recall * Precision}{Recall + Precision}$$

III. RESULTS

Table II shows how SVM and MSVM models performed. The metrics illustrate that both models performed well in all the test groups. The diseased leaves are picked out more precisely than the healthy ones. The leaves have high visual signs of disease. Healthy leaves have lower scores, indicating challenges in model performance due to less distinct visual features compared to diseased samples.

TABLE II: EVALUATION METRIC RESULTS

Metric	**Category**	**SVM**	**MSVM**
Accuracy	Diseased Leaves	89%	90%
	Healthy Leaves	72.84%	72.84%
	Overall	89%	90%
Precision	Diseased Leaves	82%	91%
	Healthy Leaves	92%	89%
Recall	Diseased Leaves	84%	95%
	Healthy Leaves	91%	80%
F1- Score	Diseased Leaves	83%	93%
	Healthy Leaves	92%	84%

The identical performance metrics imply that the MSVM model employs multiple SVM classifiers to potentially capture a broader range of features. This suggests that the complexity added by the MSVM approach translates to enhanced performance.

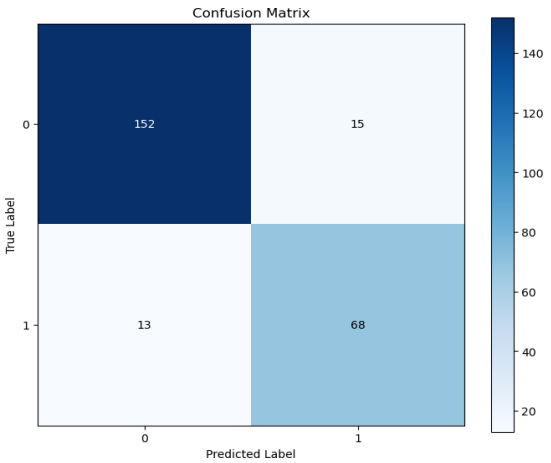

Fig 3. Confusion Matrix.

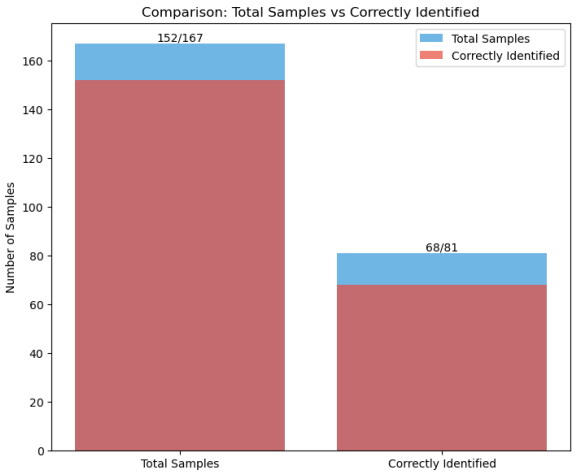

Fig 4. Comparison of Generalised Samples for SVM model.

Fig. 3 and 4 presents two visualizations related to a machine learning model\'s classification performance. The confusion matrix describes the true and false positive and false negative results, highlighting the strong performance in correctly identifying "diseased" samples with lower errors compared to "healthy" samples. This graph compares the total number of samples to the total number of correctly identified samples in each category. A higher accuracy for "diseased" samples than for "healthy" ones shows how well the model

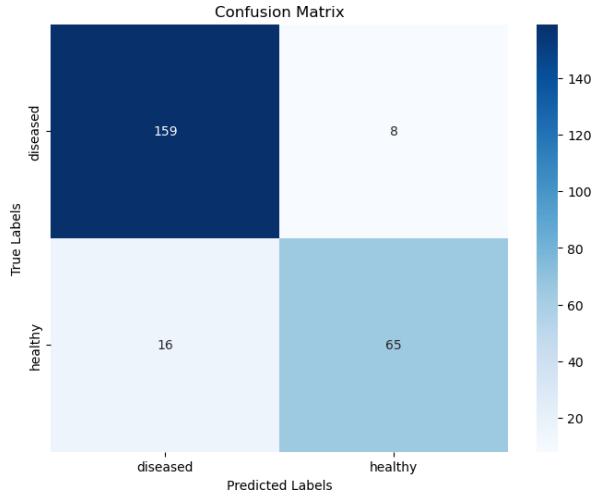

Fig 5. Distribution of generalized data samples for the MSVM model.

works

.

Fig 5 displays the number of predictions that are correct for the classes "diseased" and "healthy" for the MSVM model. This matrix clearly represents the model's ability to distinguish between the two health states of the plant samples, highlighting areas for potential improvement in classification accuracy.

Figure 6 shows how accurate different parts of the model estimation are when compared. The training accuracy is particularly high at 0.98, while the testing accuracy is 0.87. The second graph shows the overall accuracy rate (0.90) and the success rate for identifying diseased samples (0.95). This comparison shows that the model performs better in detecting diseased samples than it achieves high accuracy across the dataset, highlighting its potential advantage in disease detection over general classification.

IV. COMPARATIVE ANALYSIS

In this section, we compare our new approach with existing problem formulation frameworks. As shown in Table III , our paper shows a significant improvement in crop disease detection accuracy, reaching 90% accuracy using MSVM, which is better than the 88% accuracy of the CNN base model and the 78% accuracy of Efficient Net B-7. Accuracy. Although CNN and U-Net architectures are powerful in feature extraction and segmentation tasks, MSVM provides a more robust classification method, especially when dealing with multi-class problems in disease detection.

TABLE III: COMPARATIVE ANALYSIS TABLE

Paper	***Objective***	***Best Performing Algorithm***	***Metric Score***
[19]	Plant Disease Diagnosis Using SVM, KNN and CNN	CNN	Accuracy: 88%
[20]	Plant Disease Detection using U-Net, EfficientNet B-7	EfficientNet B-7	Accuracy: 78%
Our Paper	Crop Disease Detection Using SVM and MSVM	MSVM	Accuracy: 90%

V. CONCLUSION

A comparative analysis between SVM and MSVM for plant disease detection was performed using a dataset of plant images. By integrating GLCM for feature extraction and subsequent classification with SVM and MSVM, the models demonstrated the ability to accurately classify plant diseases. Evaluation results, including confusion matrices and performance metrics such as precision, accuracy, recall, and F1 score, indicated that both classifiers were effective, although the MSVM approach is better suited for multi-class disease classification scenarios. Overfitting is prevented through cross-validation, which ensures that the model generalizes well to unseen data. Challenges such as overfitting were observed, especially in training versus test

accuracy comparisons, indicating the need for further optimization of the model. The study highlights the effectiveness of machine learning techniques in plant disease detection and suggests that additional improvements, such as advanced data augmentation and hyperparameter tuning, can further enhance classification performance for real-world applications.

REFERENCES

[1] Albattah, W., Nawaz, M., Javed, A., Masood, M., & Albahli, S. (2021). A novel deep learning method for detection and classification of plant diseases. *Complex & Intelligent Systems, 8*(1), 507–524. https://doi.org/10.1007/s40747-021-00536-1

[2] Shoaib, M., Hussain, T., Shah, B., Ullah, I., Shah, S. M., Ali, F., & Park, S. H. (2022). Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease. *Frontiers in Plant Science, 13*. https://doi.org/10.3389/fpls.2022.1031748

[3] Saad, M. H., & Salman, A. E. (2023). A plant disease classification using one-shot learning technique with field images. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-023-17830-4

[4] Alirezazadeh, P., Schirrmann, M., & Stolzenburg, F. (2022). Improving deep learning-based plant disease classification with attention mechanism. *Gesunde Pflanzen, 75*(1), 49–59. https://doi.org/10.1007/s10343-022-00796-y

[5] Krishnan, V. G., Deepa, J. R. V. P., Rao, P. V., Divya, V., & Kaviarasan, S. (2022). An automated segmentation and classification model for banana leaf disease detection. *Journal of Applied Biology and Biotechnology, 10*(1), 213-220.

[6] Algani, Y. M. A., Caro, O. J. M., Bravo, L. M. R., Kaur, C., Ansari, M. S. A., & Bala, B. K. (2023b). Leaf disease identification and classification using optimized deep learning. *Measurement Sensors, 25*, 100643. https://doi.org/10.1016/j.measen.2022.100643

[7] Khan, H., Haq, I. U., Munsif, M., Mustaqeem, N., Khan, S. U., & Lee, M. Y. (2022). Automated wheat diseases classification framework using advanced machine learning technique. *Agriculture, 12*(8), 1226. https://doi.org/10.3390/agriculture12081226

[8] Nikith, B., Keerthan, N., Praneeth, & Amrita, T. (2023). Leaf disease detection and classification. *Procedia Computer Science, 218*, 291–300. https://doi.org/10.1016/j.procs.2023.01.011

[9] Wan, L., Li, H., Li, C., Wang, A., Yang, Y., & Wang, P. (2022). Hyperspectral sensing of plant diseases: Principle and methods. *Agronomy, 12*(6), 1451. https://doi.org/10.3390/agronomy12061451

[10] Kurmi, Y., & Gangwar, S. (2022). A leaf image localization based algorithm for different crops disease classification. *Information Processing in Agriculture, 9*(3), 456–474. https://doi.org/10.1016/j.inpa.2021.03.001

[11] Borhani, Y., Khoramdel, J., & Najafi, E. (2022). A deep learning based approach for automated plant disease classification using vision transformer. *Scientific Reports, 12*(1). https://doi.org/10.1038/s41598-022-15163-0

[12] Yadav, A., Thakur, U., Saxena, R., Pal, V., Bhateja, V., & Lin, J. C. W. (2022). AFD-Net: Apple foliar disease multi classification using deep learning on plant pathology dataset. *Plant and Soil, 477*(1–2), 595–611. https://doi.org/10.1007/s11104-022-05407-3

[13] Ahmad, A., Saraswat, D., & Gamal, A. E. (2023). A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. *Smart Agricultural Technology, 3*, 100083. https://doi.org/10.1016/j.atech.2022.100083

[14] Pandian, J., Kumar, V., Geman, O., Hnatiuc, M., Arif, M., & Kanchanadevi, K. (2022b). Plant disease detection using deep convolutional neural network. *Applied Sciences, 12*(14), 6982. https://doi.org/10.3390/app12146982

[15] Terentev, A., Dolzhenko, V., Fedotov, A., & Eremenko, D. (2022). Current state of hyperspectral remote sensing for early plant disease detection: A review. *Sensors, 22*(3), 757. https://doi.org/10.3390/s22030757

[16] Kumar, R., Chug, A., Singh, A. P., & Singh, D. (2022). A systematic analysis of machine learning and deep learning based approaches for plant leaf disease classification: A review. *Journal of Sensors, 2022*, 1–13. https://doi.org/10.1155/2022/3287561

[17] Aliyu, M. A., Mokji, M. M. M., & Sheikh, U. U. U. (2020). Machine learning for plant disease detection: An investigative comparison between support vector machine and deep learning. *IAES International Journal of Artificial Intelligence, 9*(4), 670.

[18] Sahu, S. K., & Pandey, M. (2023). An optimal hybrid multiclass SVM for plant leaf disease detection using spatial Fuzzy C-Means model. *Expert Systems with Applications, 214*, 118989.

[19] 19. Chowdhury, Muhammad EH, et al. "Automatic and reliable leaf disease detection using deep learning techniques." *AgriEngineering* 3.2 (2021): 294-312.

[20] 20. Mohameth, Faye, Chen Bingcai, and Kane Amath Sada. "Plant disease detection with deep learning and feature extraction using plant village." *Journal of Computer and Communications* 8.6 (2020): 10-22. 10.4236/jcc.2020.86002